DOCUMENT RESUME

ED 209 335 TM 810 878

AUTHOR Harnisch, Delwyn L.
TITLE Analysis of Item Response Patterns: Consistency Indices and Their Application to Criterion-Referenced Tests.
PUB DATE Apr 81
NOTE 29p.; Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Criterion Referenced Tests; Elementary Education; *Error Patterns; Junior High Schools; Quantitative Tests; *Responses; *Test Items
IDENTIFIERS Basic Arithmetic Skills Evaluation; *Individual Consistency Index; *Norm Conformity Index (Tatsuoka and Tatsuoka)

ABSTRACT

The major emphasis of this paper is in the examination of test item response patterns. Tatsuoka and Tatsuoka (1980) have developed two indices of response consistency: the norm-conformity index (NCI) and the individual consistency index (ICI). The NCI provides a measure of the degree of consistency between the response pattern of an individual and the difficulty ordering of the items for a norm group. The ICI provides a measure of the degree to which an individual's response pattern remains internally consistent within a topic or item type over time. The NCI and ICI were computed for 357, 396, and 643 third, sixth, and seventh/eighth grade students, respectively, on the Basic Arithmetic Skills Evaluation (BASE), a criterion-referenced test. There is a positive correlation between the consistency indices and the total test score across the three grade levels. The indices correlated very highly with each other. (Author/BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED209335

Analysis of Item Response Patterns: Consistency Indices and Their Application to Criterion-Referenced Tests

Delwyn L. Harnisch

University of Illinois, Urbana-Champaign

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D.L. Harnisch

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 810878

Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, April 1981.

INTRODUCTION

Traditional test theory and interpretations of test results are based on an implicit assumption that a test is equally appropriate and reliable for all examinees. Test fallibility is recognized; indeed there is much emphasis on coefficients of reliability and on errors of measurement. The errors of measurement, however, are assumed to be non-systematic, i.e., unrelated to any student characteristics. Furthermore, the variability of the errors of measurement are assumed to be constant across individuals.

The assumption that a test is equally appropriate for all individuals is a strong and an important one. In this paper, I will investigate the tenability of this assumption. It is intuitively reasonable that a criterion-referenced test may be more appropriate for some individuals than others. Unique background experiences may make an item very difficult that is quite easy for most people, e.g., the child who has never gone camping may find reading a passage about a camping experience more difficult than do children who have had the experience. Differential exposure to and emphasis of subject matter covered by a criterion-referenced test may result not only in mean differences in total score from class-to-class but also in differences in typical response patterns. Subtests of test items that are relatively difficult for most students may be relatively easy for students who have been in classes where that particular content was emphasized. Such aberrations may lead to the systematic over-or-under estimation of an individual's or group's level of achievement. That is, they tend to distort the measurement results.

The major emphasis of this research is in the examination of item response patterns. Analysis of item responses involves the evaluation of the mistakes

students made on a test. The most common student request of test administrators, counselors, homeroom or classroom teachers is: "Tell me what I'm doing wrong." (Davis, Personal communication). Present research on student errors (Brown & Burton, 1978, 1979; Davis, Jockusch & McKnight, 1978; Tatsuoka & Tatsuoka, 1980) point out that these errors are often quite systematic.

An important characteristic of error diagnosis involves the ability to tell whether and to what extent a response pattern is "typical" or "consistent." Tatsuoka and Tatsuoka (1980) speak of consistency with respect to either the average response pattern of a group or an individual's own response pattern over time. Specifically, two distinct indices have been developed by Tatsuoka and Tatsuoka (1980) to measure consistency in these two senses. Their indices are called the norm-conformity index, NCI and the individual consistency index, ICI.

The NCI provides a measure of the degree of consistency between the response pattern of an individual and the difficulty ordering of the items for a norm group. Tatsuoka and Tatsuoka (1980) have shown that the NCI is closely related to Cliff's (1977) group consistency indices. Harnisch and Linn (in press) have shown that there is a perfect negative correlation between NCI and van der Flier (1977) U' index. Sato (1981) has shown that the NCI has similar properties as his γ index. In addition, van der Flier's (1977) U' index can be shown to be the same as Sato's γ index. Furthermore, Tatsuoka and Tatsuoka (1980) have stated that by removing students with low NCI values from a data set will result in a more nearly unidimensional group data set in the sense of forming a Guttman scale.

The ICI provides a measure of the degree to which an individual's response pattern remains internally consistent within a topic or item type overtime. For example, in a criterion-referenced test consisting of three parallel

subtests, the ICI indicates whether an examinee's response pattern changes markedly from one subtest to the next or remains relatively stable. An individual with a low ICI suggests that the examinee was in the formative stages of learning, alternating methods for solving equivalent problems from one subtest to the next. An individual with a high ICI value indicates the nearing of mastery or a learning plateau which is reflected in his stable response pattern.

A criterion-referenced test is composed of items measuring specific skill areas. Most criterion-referenced tests have at least three items measuring each skill. The items correct for each skill yield a specific skill score. The NCI requires that the skills are ordered in descending order of difficulty for a particular group, the norm group. The NCI indicates the degree of consistency between the response pattern, $\underset{\sim}{S}$, and the difficulty ordering of the skills for a norm group. For example, the following matrix, $\underset{\sim}{S}$, represents the student data matrix for 4 skills each measured by 3 items:

$$\underset{\sim}{S} = \begin{vmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{vmatrix}$$

Next, we determine the transpose of the complement of $\underset{\sim}{S}$ denoted as $\overline{\underset{\sim}{S}}'$:

$$\overline{\underset{\sim}{S}}' = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{vmatrix}$$

The product of the transpose of the complement, $\overline{\underset{\sim}{S}}'$, and the student data matrix, $\underset{\sim}{S}$, yields a dominance matrix, $\underset{\sim}{N}$:

$$\underset{\sim}{N} = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{vmatrix} \begin{vmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 0 & 1 & 2 & 3 \\ 0 & 0 & 2 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{vmatrix}$$

For the elements, n_{ij}, of the dominance matrix, $\underset{\sim}{N}$, two terms are defined which will be used in determining the NCI for an individual. The first term is designated at U_a which is the sum of the upper triangle elements of $\underset{\sim}{N}$ and defined as:

$$U_a = \sum_{i=1}^{I} \sum_{j=i+1}^{J} n_{ij}$$

The second term is designated as U which is the sum of all the elements of $\underset{\sim}{N}$ and defined as:

$$U = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij}$$

The NCI for an individual may now be expressed in terms of U_a and U as being:

$$NCI = \frac{2U_a}{U} - 1$$

Thus, the NCI for an individual whose response matrix was given above would have an NCI = $\frac{2(11)}{12} - 1 = .83$. NCI's may range from -1 to 1.0. A value of 1.0 indicates perfect consistency with the norm-group ordering of skill difficulties, that is, correct responses to the $n_{i.}$ easiest skill items and incorrect responses to the remaining $J-n_{i.}$ skill items.

The ICI is also based on a dominance matrix but only after the skill scores are ordered based on the examinee's response pattern versus the norm group. The ICI for an individual is computed in the same manner as the above NCI once the skills are arranged in descending order of difficulty. For example, suppose an individual took a test covering 4 skills with 3 items on each skill and had a response matrix, $\underset{\sim}{S}$, as shown below where the skills are arranged according to

difficulty based on a norm group.

$$\underset{\sim}{S} = \begin{vmatrix} 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{vmatrix}$$

First, we rearrange the skills in $\underset{\sim}{S}$ to reflect the descending order of difficulty for this individual. Rearranged $\underset{\sim}{S}$ is:

$$\underset{\sim}{S} = \begin{vmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{vmatrix}$$

Thus, the transpose of the complement of rearranged $\underset{\sim}{S}$, $\overline{\underset{\sim}{S}}'$ is shown below as:

$$\overline{\underset{\sim}{S}}' = \begin{vmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix}$$

Thus, the dominance matrix, $\underset{\sim}{N}$ for this individual is:

$$\underset{\sim}{N} = \overline{\underset{\sim}{S}}'\,\underset{\sim}{S} = \begin{vmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix} \begin{vmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{vmatrix} = \begin{vmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{vmatrix}$$

The ICI based on the above dominance matrix would have a value = $\frac{2(8)}{9} - 1 = .78$. The NCI for this same individual would have a dominance matrix equal to:

$$\underset{\sim}{N} = \overline{\underset{\sim}{S}}'\underset{\sim}{S} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{vmatrix} \begin{vmatrix} 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{vmatrix} = \begin{vmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 \\ 1 & 2 & 1 & 0 \end{vmatrix}$$

Thus, the NCI for this individual would have a value = $\frac{2(2)}{9} - 1 = -.56$. The above example illustrates that ICI $\geq$ NCI. This results since the skills are arranged in their order of difficulty based on the individual while for the computation of NCI the skills are arranged in their order of difficulty for a norm group. Tatsuoka and Tatsuoka (1980) have shown that the following property exists between ICI and NCI, namely that ICI $\geq$ NCI for each individual.

When an individual's test score was either zero or perfect, a value of 1 was assigned to their respective values of NCI and ICI. Interrelationships of NCI and ICI to each other and to other variables are reported below. Evaluation of the NCI and ICI in combination with the total test score will be shown for each of the data sets below.

Results and Discussion

The NCI and ICI were computed for 357, 396, and 643 3rd, 6th, and 7th-8th grade students, respectively, on the Basic Arithmetic Skills Evaluation (BASE) testing system. (See Appendix A for a copy of the sample items used at each of the 3 grade levels.) BASE is a criterion-referenced test with 3 items per skill area. The number of skills assessed at grades 3, 6, and 7-8 were 23, 21, and 40, respectively. In addition, measures of student background, teacher and school characteristics were available at each grade.

The descriptive statistics of the NCI, ICI, total test score, and total skills mastered were then computed for each test at each of the three grade levels. The quartile score points, mean and standard deviation are given for the NCI and ICI at each of the three grade levels in Table 1.

- - - - - - - - - - - - - -

Insert Table 1 about here

- - - - - - - - - - - - - -

The quartile score points of the NCI and ICI for the 3rd and 7th-8th grade levels had very similar values while the quartile score points for the 6th grade level were much lower. The lower quartile score points on the NCI at grade 6 suggests that the ordering of skills based on the norm group is not as consistent with the response pattern for these students. This result suggests that the ordering of the skills is not the same for the individual student and moreover this lower consistency index may be due to differential exposure to and emphasis of skills covered by the BASE test. Furthermore, the difference between the NCI and ICI reveals that the students response pattern is quite consistent over the skills when skills are arranged by the individual's ordering of skills suggesting even further that these differences may be due to teacher coverage of and emphasis of subject matter. Evaluation of the NCI and ICI at the teacher level will be investigated and a report of the results will be shown below.

Scatterplots of the indices with each other were made and are shown in Figure 1 for each of the three grade levels.

- - - - - - - - - - - - - - -

Insert Figure 1 about here

- - - - - - - - - - - - - - -

Since the ICI $\geq$ NCI, all of the points in the scatterplot fall below the major diagonal. The consistency indices may be used along with the total test score to identify types of students who need more study, who make careless mistakes, who possess sporadic study habits or insufficient readiness, or who are doing everything fine. Table 2 provides a profile of student status based on combinations of high and low NCI and ICI values along with the total test score earned.

- - - - - - - - - - - - - - -

Insert Table 2 about here

- - - - - - - - - - - - - - -

Students with a high total test score and possessing both a high NCI and a high ICI value are students who need no more study since all is well. However, students with a low total test score and possessing both a high NCI and a high ICI value are students who are making consistent mistakes by perhaps using an incorrect rule which yields correct answers to easy problems and incorrect answers to hard problems.

Students with both a low NCI and a low ICI value are students whose mistakes or errors made is in a random pattern. If cutting scores are the same on the NCI and ICI, no students will be identified in the cell labelled with a high NCI and a low ICI value since all the students have ICI values greater than or equal to their respective NCI value.

Students with a high total test score and possessing both a low NCI and a high ICI value are students making careless mistakes on a few of the easy problems. However, the students with a low total test score and possessing both a low NCI and a high ICI value are students needing more study since they are getting many easy problems wrong.

Scatterplots of the NCI with the total test score are shown in Figure 2 for each of the three grade levels. As shown in Table 2, the NCI and ICI are useful in combination with total test score for diagnosing student's problems. Therefore, a low and high ability group of examinees was formed at each grade level and scatterplots were made for NCI by ICI as shown in Figure 3 only for the grade 7-8 students.

- - - - - - - - - - - - - - - - - -

Insert Figures 2 and 3 about here

- - - - - - - - - - - - - - - - - -

Examination of the scatterplots in Figure 3 reveals the number of students possessing certain types of problems and furthermore allows us to compare the seriousness of problems across the three grade levels.

The indices were then correlated with each other and with other selected variables including total test score. These correlations were computed separately at each of the three grade levels and are shown in Table 3.

- - - - - - - - - - - - - -

Insert Table 3 about here

- - - - - - - - - - - - - -

The correlations of NCI with ICI across the 3 grade levels ranged from .714 to .788. The correlation of NCI with the total test score was highest with grade 6 students at .412 while the correlation for 3rd and 7th-8th grade students was respectively, .156 and .147. The correlations of ICI with total test score revealed a similar pattern as did the correlation of NCI with the total test score.

The scatterplots of NCI versus total test score as shown in Figure 2 reveal the effects of outliers on the correlation between NCI and total test score for students in the lower and upper grades. Each of these plots has high ability level students making some careless mistakes thus resulting in their high ICI but low NCI index.

Further analyses were performed with each grade level data. An analysis of variance was performed for NCI and ICI each as dependent variables and teacher and race of student as the two independent factors. Significant main effects were found with both dependent variables, NCI and ICI, on the teacher factor for students at the grade 7-8 level. For students at the grade 6 level, significant main effects were found with the teacher factor on only the NCI index. The ANOVA results are summarized in Table 4.

- - - - - - - - - - - - - -

Insert Table 4 about here

- - - - - - - - - - - - - -

The differences on the NCI's for both the grade 6 and grade 7-8 students with different teachers ranged from .326 to .707 suggesting that these differences may have resulted from differential treatment to the topics covered on the test.

Since the total test score was correlated with both consistency indices, an ANCOVA was performed with the same dependent and independent variables. The results are summarized in Table 5. Significant main effects remained only with grade 7-8 students from different classrooms suggesting that the differential treatment to the topics covered on the test is worth investigating further to determine which subset of items were contributing the greatest to this teacher effect.

- - - - - - - - - - - - - -

Insert Table 5 about here

- - - - - - - - - - - - - -

Of additional interest would be to recompute the NCI's for students whose mean value at the teacher level indicated unusual response patterns using the skill difficulty ordering determined by individual teachers and/or clusters of teachers. These analyses would possibly provide an evaluation of the extent to which students who appear to have unusual response patterns in comparison to district data are actually relatively consistent with results for their own classroom or identifiable clusters of classrooms. Such an outcome would suggest the need for teacher, school, and program comparisons to be made on subsets of items rather than only on total scores.

In summary, there is positive correlation between the consistency indices and the total test score across the 3 grade levels. The indices correlate very high with each other ranging in value from .714 to .788.

The scatterplots of the NCI by total test score and NCI by ICI for all students as well as for low and high ability students further reveals the nature of the relationship between the indices and the total test score. Forming low and high ability groups based on their total test performance for plots of NCI vs ICI advances our understanding of student performance on criterion-referenced tests. The teacher differences were evident in the differences for students on both the NCI and ICI. Controlling for total test score removed the teacher effect from two of the three grade levels studied. The significant teacher effect for grade 7-8 students after controlling for total test score is under additional investigation to determine which subset of items is contributing to the identification of students with unusual response patterns.

References

Brown, J. S., & Burton, R. B. Diganostic models for procedural bugs in basic mathematical skills. Cognitive Science, 1978, 2, 155-192.

Brown, J. S., & Burton, R. B. Diagnostic models in basic mathematical skills. In Testing, teaching and learning. Report of a conference on research on testing. Washington, D.C.: National Institute of Education, 1979, 273-310.

Cliff, N. A theory of consistency of ordering generalizable to tailored testing. Psychometrika, 1977, 42, 375-399.

Davis, R. B. Personal communication, March 1981.

Davis, R. B., Jockusch, E., & McKnight, C. Cognitive processes in learning algebra. The Journal of Children's Mathematical Behavior, 1978, 2, 1-320.

Harnisch, D. L., & Linn, R. L. Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. Journal of Educational Measurement, in press.

Sato, T. Similarity of some indices of item response patterns (Research Memorandum EI81-1). Tokyo: Computer & Communications Systems Research Laboratories, Nippon Electric Co., Ltd., 1981.

Tatsuoka, M., & Tatsuoka K. Detection of aberrant response patterns and their effect on dimensionality (Research Report 80-4). Urbana: University of Illinois, Computer-based Education Research Laboratory, 1980.

van der Flier, H. Environmental factors and deviant response patterns. In Y. H. Poortinga (Ed.), Basic problems in cross-cultural psychology. Amsterdam: Swets and Seitlinger, B.V., 1977.

Table 1

Quartile Score Points, Means and Standard Deviations on the NCI and ICI for Students at Three Grade Levels

			Quartile				Standard
Index	n	Grade	1st	2nd	3rd	Mean	Deviation
NCI	357	3	.419	.706	1.000	.605	.400
	396	6	.385	.630	.906	.602	.339
	643	7-8	.467	.667	.818	.620	.268
ICI	357	3	.85[illegible]	1.000	1.000	.896	.183
	396	6	.789	.905	1.000	.867	.150
	643	7-8	.800	.889	1.000	.868	.141

Table 2

Profile of Students with High and Low NCI, ICI, and Total Test Score

NCI \ ICI	Low	High
High	No students in this cell if cutting scores are the same on NCI and ICI since ICI $\geq$ NCI always.	If total test score is low, student is making consistent mistakes by perhaps using an incorrect rule which yields correct answers to easy problems and incorrect answers to hard problems. If total test score is high the student needs no more study since all is well.
Low	Students with a random error pattern.	If total test score is low, the student is getting many easy problems wrong. The student needs more study. If total test score is high, the student is making careless mistakes on a few of the easy problems.

Table 3

Correlations Between Consistency Indices and Total Test Score for Students at Three Grade Levels

Variable	n	Grade	NCI	ICI
ICI	357	3	.788	
	396	6	.714	
	643	7-8	.747	
Total Test Score	357	3	.156	.163
	396	6	.412	.326
	643	7-8	.147	.125

Table 4

Analysis of Variance Summary Results for NCI and ICI

with Grade 3, 6, and 7-8 Students

	Grade 3			Grade 6			Grade 7-8		
Source	df	Mean Square	F	df	Mean Square	F	df	Mean Square	F
				NCI					
Main Effects									
Teacher	17	.208	1.345	13	.201	1.810*	9	.219	3.522**
Race	1	.021	.135	1	.266	2.397	1	.091	1.465
Explained	18	.198	1.282	14	.208	1.873*	10	.216	3.483**
Residual	326	.155		363	.111		605	.062	
				ICI					
Main Effects									
Teacher	17	.043	1.325	13	.030	1.377	9	.058	3.252**
Race	1	.006	.169	1	.074	3.354	1	.041	2.291
Explained	18	.042	1.277	14	.035	1.577	10	.058	3.262**
Residual	326	.033		363	.022		605	.018	

*p < .05

**p < .01

Table 5

Analysis of Covariance Summary Results for NCI and ICI with Total Test Score as Covariate for Grade 3, 6, and 7-8 Students

Source	Grade 3 df	Grade 3 Mean Square	Grade 3 F	Grade 6 df	Grade 6 Mean Square	Grade 6 F	Grade 7-8 df	Grade 7-8 Mean Square	Grade 7-8 F
NCI									
Covariate									
Total Test Score	1	1.466	9.682**	1	6.702	69.500**	1	.819	13.401**
Main Effects									
Teacher	17	.191	1.262	13	.115	1.191	9	.185	3.031**
Race	1	.018	.116	1	.007	.070	1	.003	.048
Explained	19	.249	1.646*	15	.548	5.681**	11	.226	3.704**
Residual	325	.151		362	.096		604	.061	
ICI									
Covariate									
Total Test Score	1	.335	10.501**	1	.761	37.213**	1	.106	5.980*
Main Effects									
Teacher	17	.043	1.355	13	.024	1.182	9	.054	3.045**
Race	1	.005	.152	1	.005	.230	1	.015	.844
Explained	19	.056	1.767	15	.072	3.523**	11	.055	3.106**
Residual	325	.032		362	.020		604	.018	

*p < .05

**p < .01

Figure Captions

Figure 1. NCI vs ICI for each grade Level.

Figure 2. Scatterplots of NCI by total test score for each of the three grade levels.

Figure 3. Scatterplots of NCI by ICI for students of low and high ability at the grade 7-8 level.

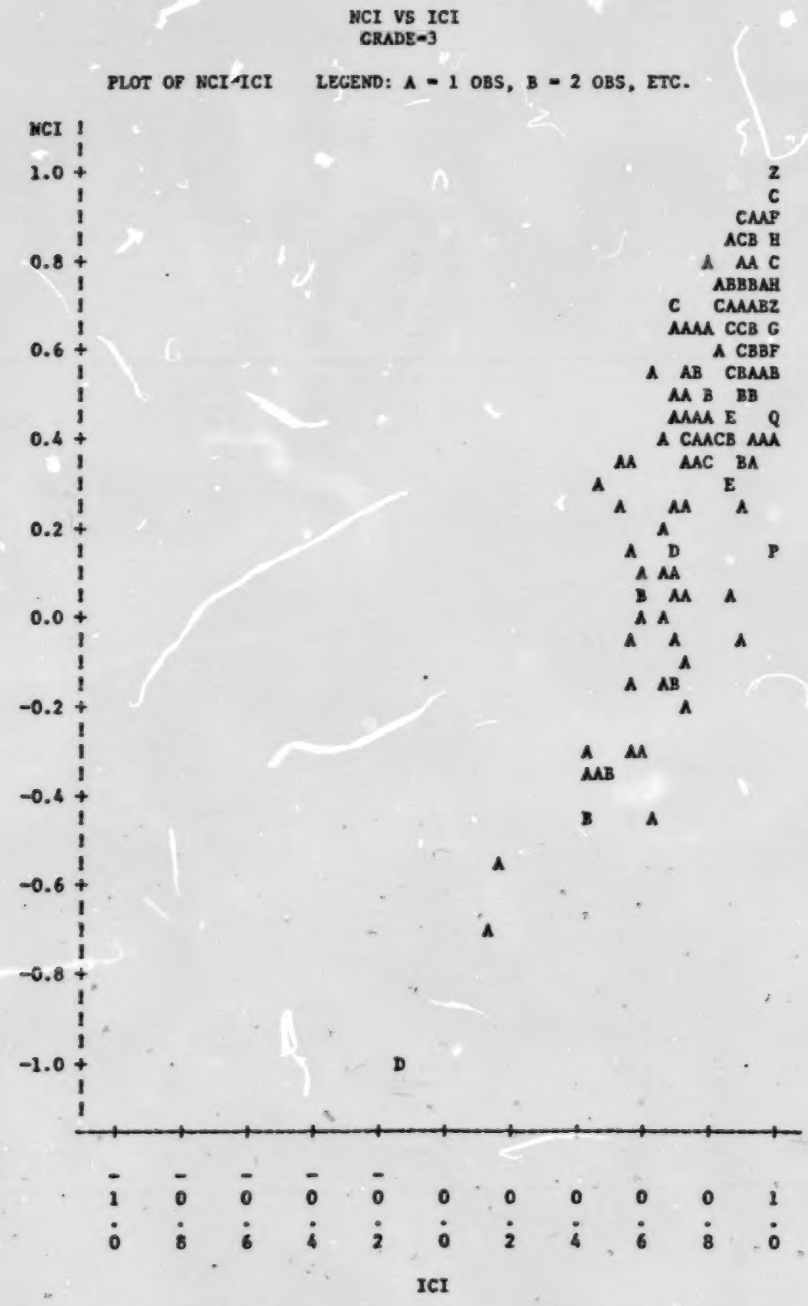


Figure 1. NCI vs ICI for each grade level.

Figure 1 (con't.)

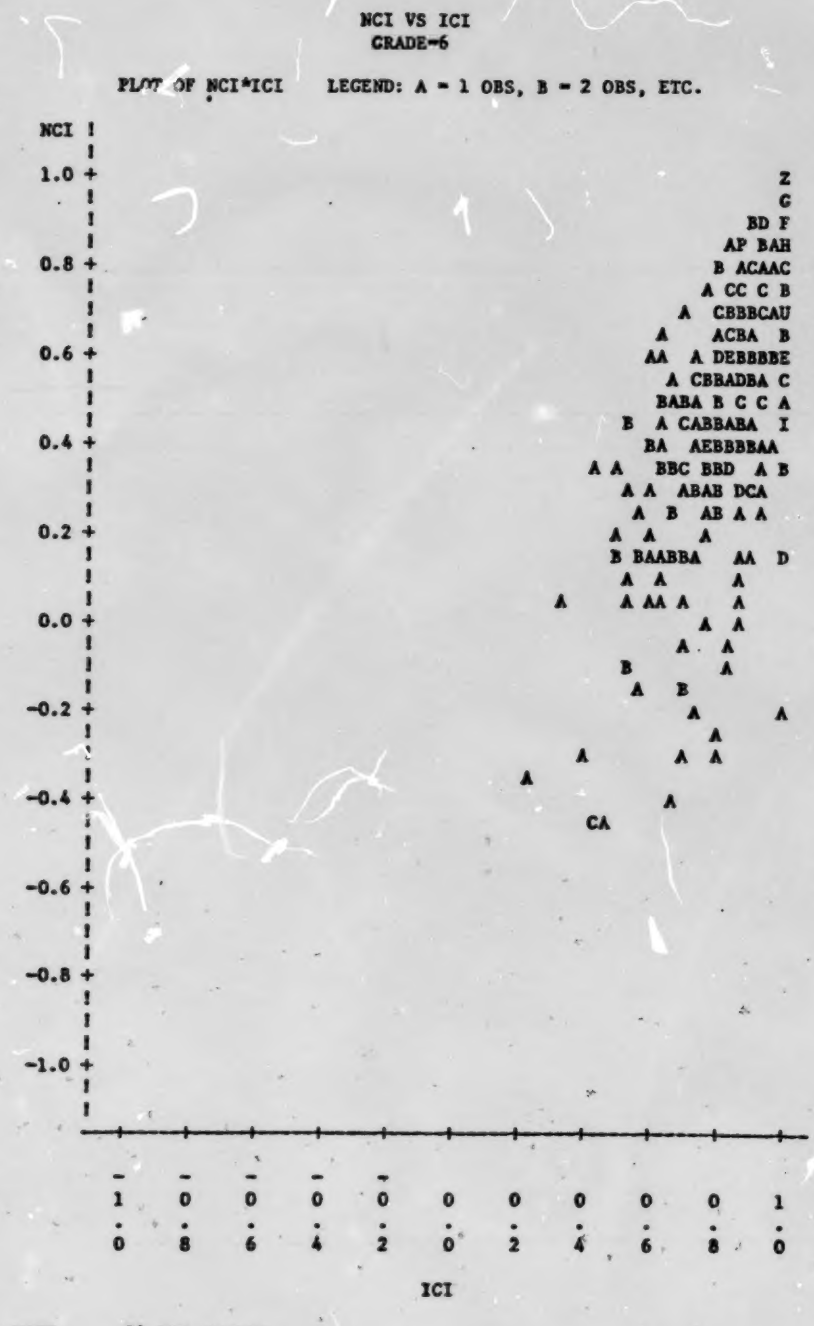
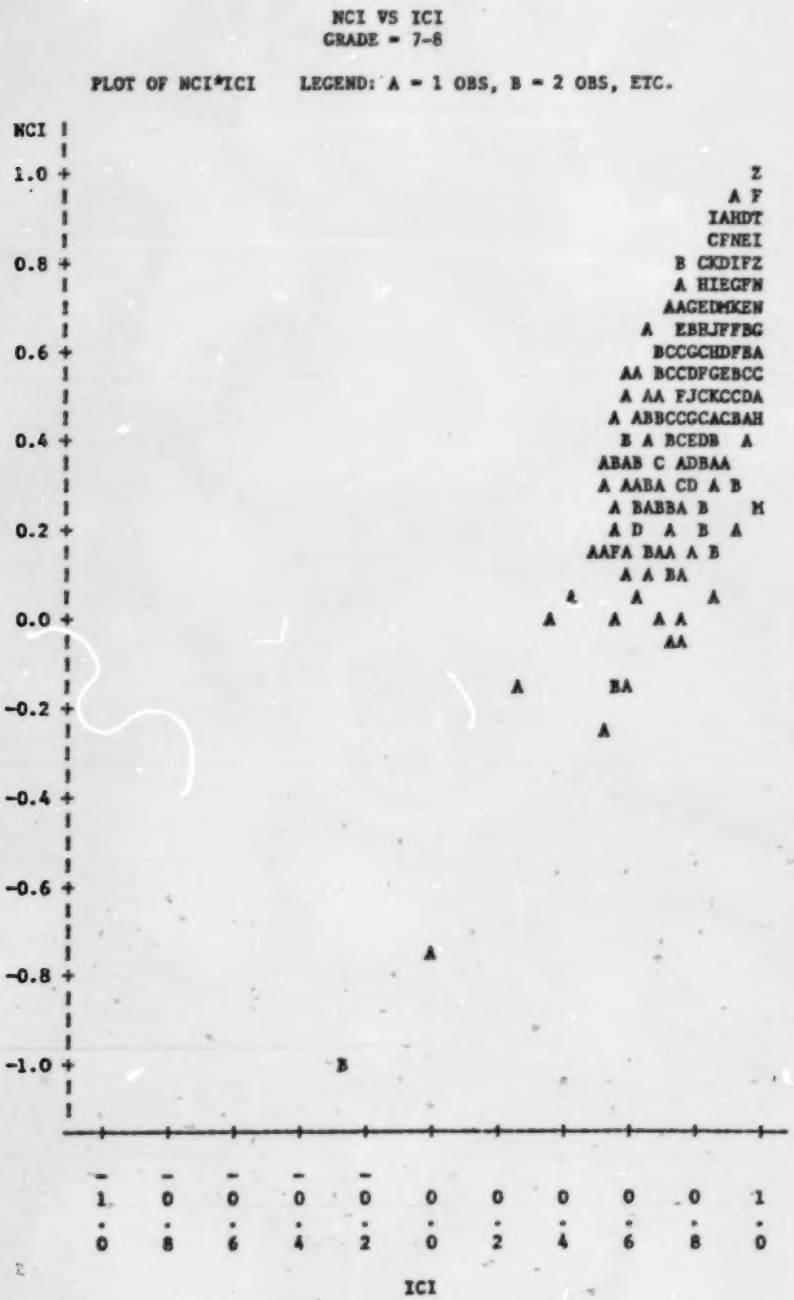


Figure 1 (con't.)



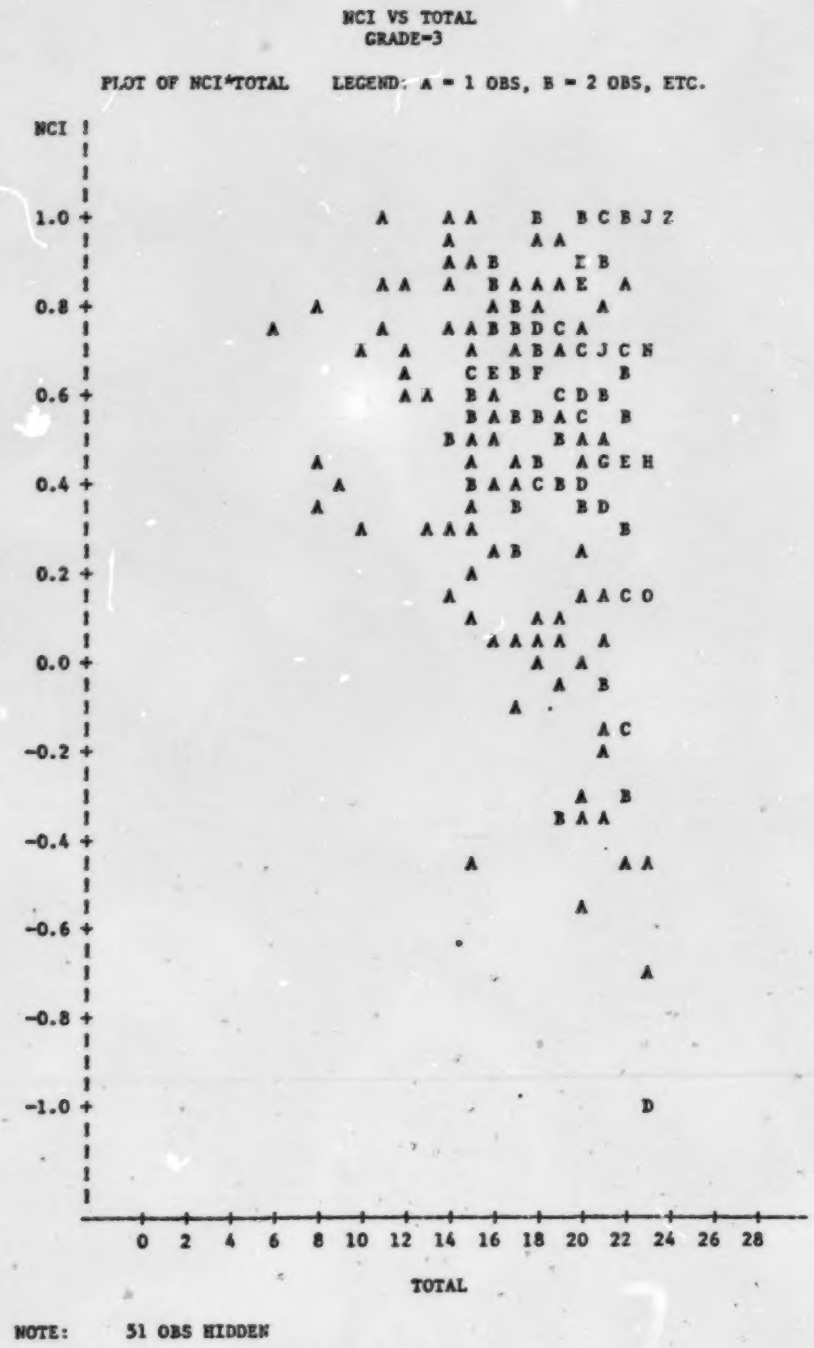
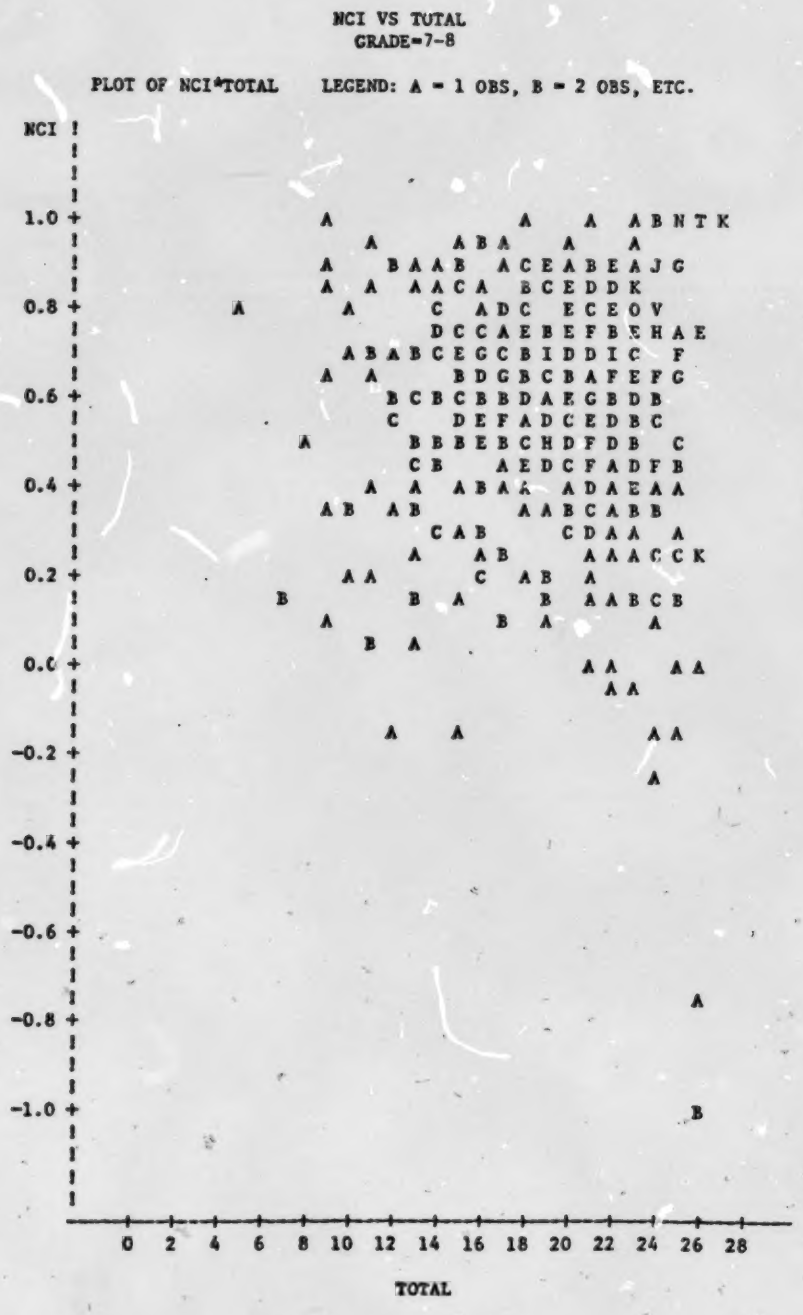


Figure 2. Scatterplots of NCI by total test score for each of the three grade levels.

Figure 2 (con't.)

Figure 2 (con't.)



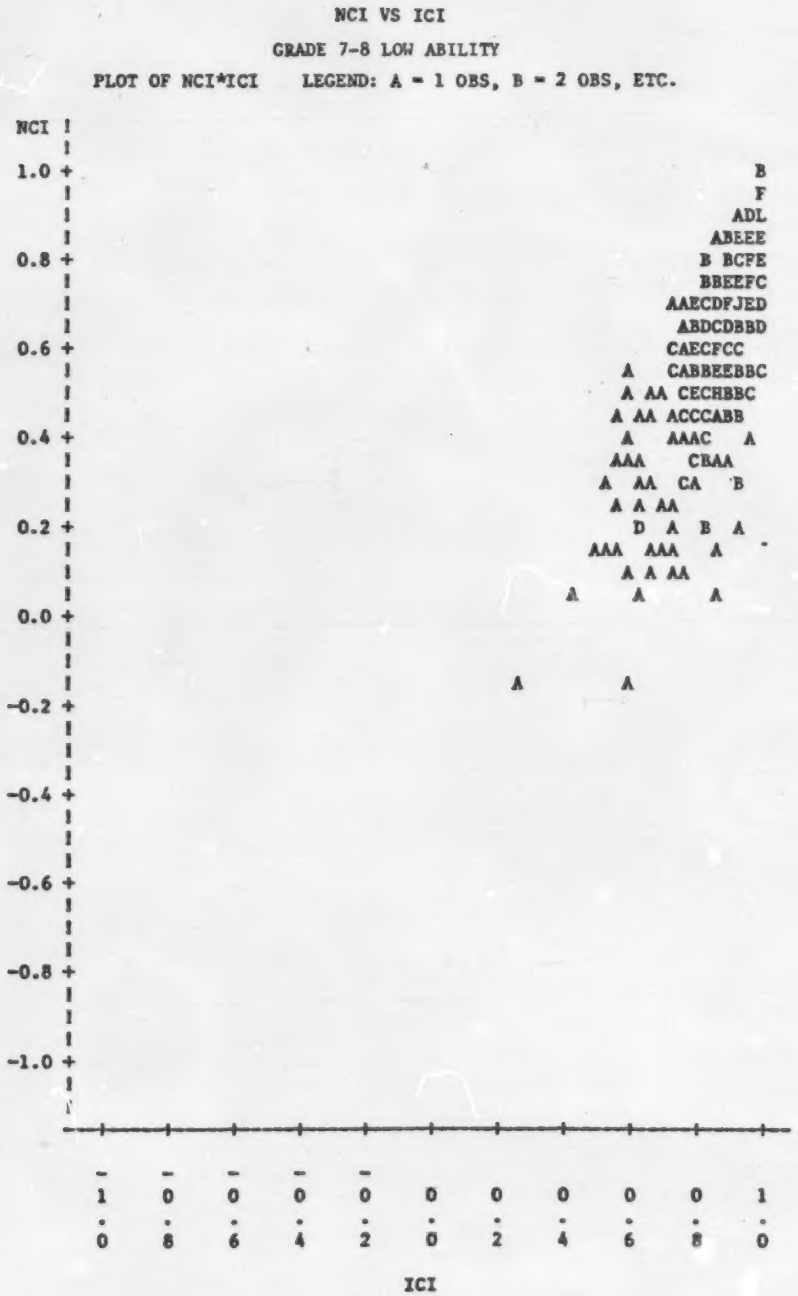
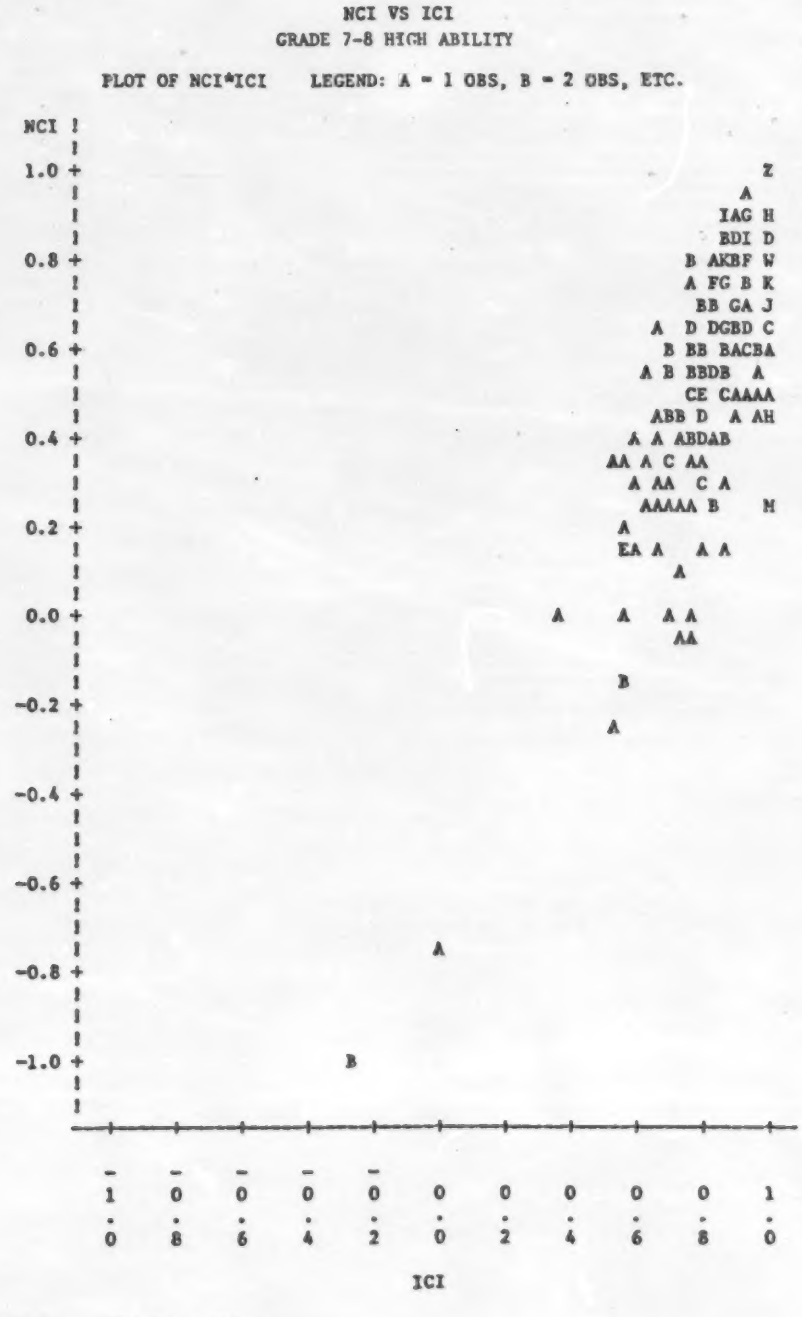


Figure 3. Scatterplots of NCI by ICI for students of low and high ability at grades 7-8.

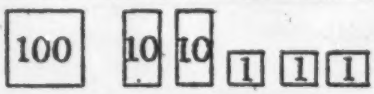
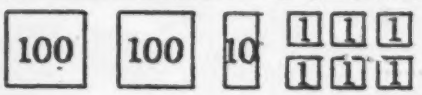
Figure 3 (con't.)



Appendix A

BASE Sample Items

Sample Items From BASE for Grade 3

①	②	③
What number is shown? 	What number is shown? 	What number is shown?
123 231 312	126 261 216	332 233 323 STOP

Sample Items From BASE for Grade 6

①	②	③
What is the place value of digit 7?	What is the place value of digit 8?	What is the place value of digit 2?
867,423,028	893,765,412	923,654,780
7,000,000 700,000 7,000	800,000,000 800,000 800	20 20,000,000 20,000 STOP

Sample Items From BASE for Grades 7-8

①	②	③
4,789 +13,678	25,069 + 8,975	35,679 +44,827
18,466 18,467 17,457	33,044 33,944 34,044	80,506 80,406 80,496